View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Developmental Biology 288 (2005) 595 – 611

DEVELOPMENTAL BIOLOGY

www.elsevier.com/locate/ydbio

Genomes & Developmental Control

Genome-wide expression dynamics during mouse embryonic development reveal similarities to *Drosophila* development

Roger A. Wagner *, Raymond Tabibiazar, Arnold Liao, Thomas Quertermous

Division of Cardiovascular Medicine, Stanford University School of Medicine, Falk CVRC, 300 Pasteur Drive, Stanford, CA 94305, USA

Received for publication 2 March 2005, revised 16 July 2005, accepted 1 September 2005

Abstract

Gene transcription mediates many vital aspects of mammalian embryonic development. A comprehensive characterization and analysis of the dynamics of gene transcription in the embryo is therefore likely to provide significant insights into the basic mechanisms of this process. We used microarrays to map transcription in the mouse embryo in the important period from embryonic day 8 (e8.0) to postnatal day 1 (p1) during which the bulk of the differentiation and development of organ systems takes place. Analysis of these expression profiles revealed distinct patterns of gene expression which correlate with the differentiation of organs including the nervous system, liver, skin, lungs, and digestive system, among others. Statistical analysis of the data based on Gene Ontology (GO) group annotation showed that specific temporal sequence patterns in gene class utilization across development are very similar to patterns seen during the embryonic development of *Drosophila*, suggesting conservation of the temporal progression of these processes across 550 million years of evolution. The temporal profiles of gene expression and activation of processes revealed here provide intriguing insights into the mechanisms of mammalian development, embryogenesis, and organogenesis, as well as into the evolution of developmental processes.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Development; Embryogenesis; Organogenesis; Transcriptional profiling; Evolutionary biology

Introduction

Mammalian embryonic development is a complex interactive dance which involves cell–cell signaling, cell migration, differentiation, apoptosis, and a host of other biological processes. Key elements of each of these processes are controlled at the transcriptional level, and identification of the genes involved in particular processes is often the first step toward analysis of their functions at the biochemical and cell biological level. Elegant studies of gene transcription throughout embryogenesis and the life cycle of *Drosophila* and *C. elegans* have shown the power that organized analysis of transcription on the genome scale can have in revealing insights into the organization of these processes and the genes involved in them (Arbeitman et al., 2002; Birnbaum et al., 2003; Furlong et al., 2001; Hill et al., 2000; Jiang et al., 2001; Kim et al., 2001).

The mouse is the primary model for the study of mammalian development and disease because of its size, short generation time, and the ease of creation of gene knockout mutants. Mouse embryonic development has been extensively characterized at the morphologic level, and the time of onset of many developmental processes is known precisely. However, the transcriptional background upon which these processes are based has been systematically studied at a genome wide level only during preimplantation development (Hamatani et al., 2004; Wang et al., 2004; Zeng et al., 2004). Many of the most important events in terminal differentiation and organogenesis occur significantly later in development, between embryonic days e8.0 and birth, and include differentiation and maturation of somites and their mesodermal progeny as well as morphogenesis and maturation of the heart and circulatory system, nervous system, digestive system, and organs such as the liver, kidney, lung, skin, and pancreas. Gaining a broader knowledge base about the genes involved in these processes will allow us to focus on those which are utilized in normal development, and which, if misexpressed, may lead to altered development and disease.

* Corresponding author.
E-mail address: rwagner@cvmed.stanford.edu (R.A. Wagner).

0012-1606/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2005.09.036

Microarray analysis is currently the most powerful tool for assessing global gene expression patterns. We have developed mouse cDNA microarrays optimized for the analysis of embryonic development by combining genes from the NIA 15K mouse developmental gene set (Carter et al., 2003), composed primarily of genes expressed in early embryonic development, with the Riken 22K mouse gene set (Miki et al., 2001) derived from older embryonic and adult tissues. These were supplemented with several thousand ES cell line derived ESTs and known developmentally important genes to create arrays containing 43,200 gene features representing ~25,000 unique genes. We have used these microarrays to systematically assess the global gene expression profile of mouse embryos from e8.0 to postnatal day 1 when much of organogenesis occurs.

Materials and methods

Embryo collection

Details are available as Supplemental data. Embryos were obtained from timed pregnancies of C57Bl/6 mice and staged by careful observation of morphological criteria (Kaufman and Bard, 1999) to obtain three separate pools of 10–15 embryos for each of the 18 developmental stages studied, e9.0, e9.5, e10.0, e10.5 (n = 15/pool), and e11.0, e11.5, e12.0, e12.5, e13.0, e13.5, e14.5, e15.0, e15.5, e16.5, e17.5, e18.5, and P1 (n = 10/pool), with the exception of e8.0 (n = 60 embryos in 1 pool).

RNA extraction, labeling, and microarray hybridization

Details are available as Supplemental data. RNA was extracted from each pool of embryos using Trizol extraction followed by purification on Qiagen RNeasy columns (Qiagen, Inc). A universal mouse reference RNA was prepared from e17.5 mouse embryos and used for all hybridizations so that all arrays could be compared (Miki et al., 2001). Three biological replicate microarray hybridizations using 30 μg of input RNA to create labeled Cy3 (timepoint) or Cy5 (common reference) cDNA were performed for each embryonic stage (except e8.0, n = 1) and hybridized as previously described (Tabibiazar et al., 2003), for a total of 2,203,200 gene expression measurements.

Data analysis

Details are available as Supplemental data. Microarrays were scanned using an Agilent G2565AA scanner, and the data were LOWESS normalized using GeneData Refiner software (GeneData Corporation, South San Francisco, CA). The data were analyzed using GeneData Expressionist microarray analysis tools, Significance Analysis of Microarrays (Tusher et al., 2001), Hi-Throughput GOMiner (Zeeberg et al., 2003), Heatmap Builder (Ashley et al., 2004), and TIGR Multiple Experiment Viewer (Saeed et al., 2003) Data from the e8.0 timepoint were not included in comparative statistical analyses because this timepoint was represented by only a single microarray and because embryos at this timepoint were not dissected away from yolk sac, making comparisons with other samples derived only from embryo problematic. Some genes are represented more than once on the array, and different ESTs may correspond to the same gene. Thus, while it would be most accurate to refer to "array elements," for simplicity, we refer to them as "genes" instead. Primary data and annotation are available on our website (http://www.mousedevelopment.org) and at ArrayExpress (http://www.ebi.ac.uk/arrayexpress/).

Results and discussion

Stage-specific gene expression profiles during normal mouse embryogenesis

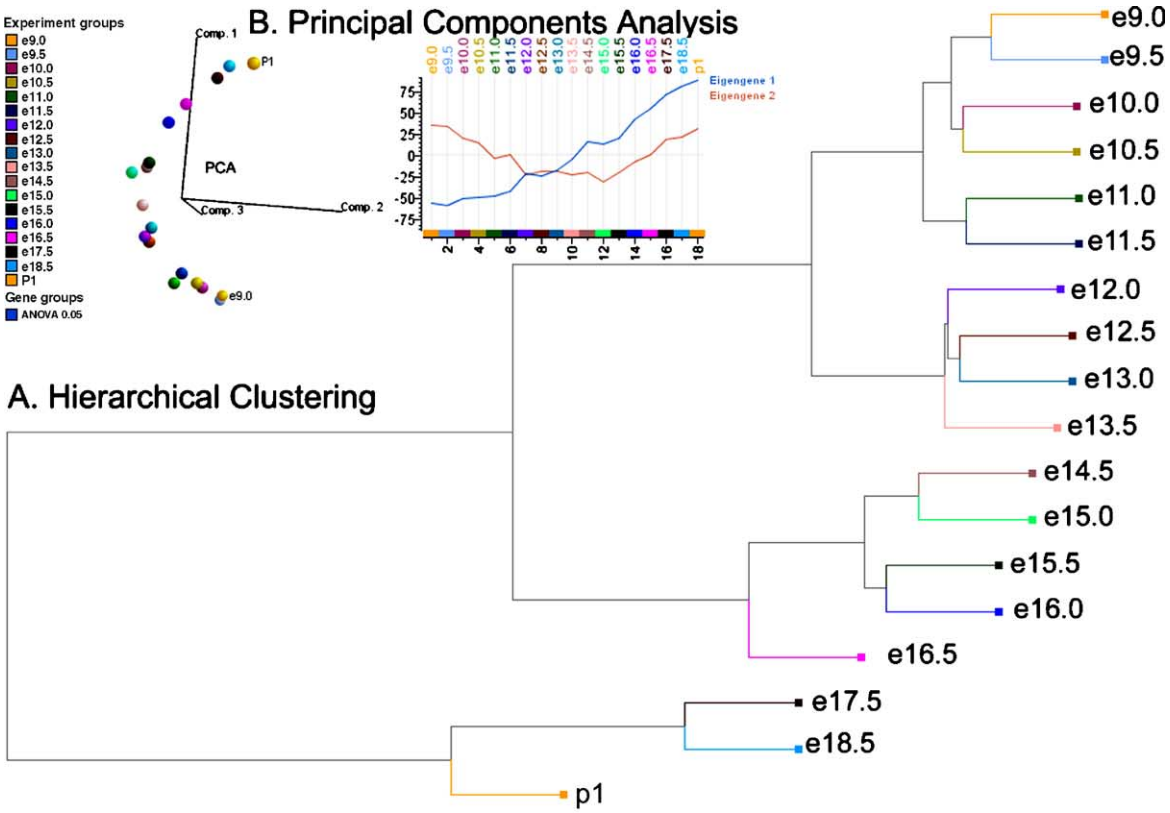
RNA was harvested from three separate pools of 10–15 embryos for each of the 18 developmental stages studied, e8.0, e9.0, e9.5, e10.0, e10.5, e11.0, e11.5, e12.0, e12.5, e13.0, e13.5, e14.5, e15.0, e15.5, e16.5, e17.5, e18.5, and postnatal day 1 to obtain proper biological and technical replicates for probe labeling and array hybridization. Pooling minimizes technical and statistical issues resulting from embryo-to-embryo variation in gene expression (Agrawal et al., 2002; Churchill, 2002; Kendziorski et al., 2003; Peng et al., 2003). Embryos were staged by careful observation of morphological criteria (Kaufman and Bard, 1999). A universal mouse reference RNA was prepared from e17.5 mouse embryos and used for all hybridizations so that all arrays and experiments could be compared (Miki et al., 2001). The e17.5 embryo reference RNA resulted in a signal of at least 2 standard deviations above background, allowing valid measurements to be made, for ~93–95% of array elements. Three biological replicate microarray hybridizations were performed for each embryonic stage (except e8.0, n = 1), as previously described (Tabibiazar et al., 2003; Wagner et al., 2004), for a total of 2,203,200 gene expression measurements, with an average correlation between biological replicates of 0.77 to 0.90. Annotated expression profiles for any gene can be interactively accessed by gene name, accession number, or BLAST search at www.mousedevelopment.org. (Supplementary data and tables associated with this article can be found in the online version or at http://microarray-pubs.stanford.edu/mousedevelopment/).

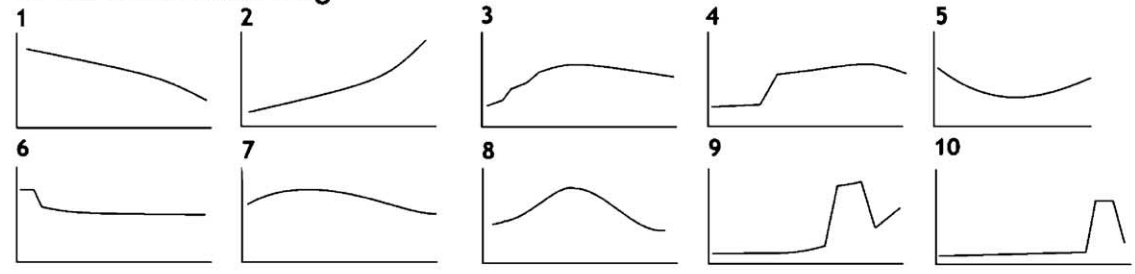
Global overview of gene expression in the embryo

We sought to establish a global overview of gene expression patterns in the embryo by using three types of statistical analyses, hierarchical clustering, principal components analysis (PCA), and K-means clustering. Initially, a group of 17,491 variable genes from e9.0 to P1 was defined using analysis of variance (ANOVA, $P < 0.005$) and was used for these subsequent analyses. The e8.0 timepoint was not included in statistical analysis because only one array hybridization could be performed due to limited RNA availability.

The timepoints were compared using hierarchical clustering to determine which were the most similar in terms of gene transcription (Sherlock, 2000). Hierarchical clustering demonstrates that adjacent developmental stages have very similar but distinct expression profiles, with increasing distance between nodes as difference in stage increases (Fig. 1A). This suggests

Fig. 1. Global overview of transcription during embryonic development. (A) Hierarchical clustering of timepoints demonstrates an orderly progression in transcription across the time course. (B) Principal Components Analysis identifies the major trends in the data, characterized as Eigengenes. Note the predominant theme of gradual changes in gene regulation across time. (C) K-means clusters — schematic representation of predominant cluster patterns. See Supplemental Tables for full lists of genes included in each cluster. (D) Representative genes contributing to major GO biological process groups overrepresented in selected K-means clusters.





D.

Cluster 3 Neurogenesis			Cluster 4 Heme Metab/Erythrocyte	Cluster 5 Glycolysis/Sugar Metab	Cluster 9 Epidermal Diff'n/Keratinocyte	Cluster 10 Digestion/lipid/protein	
ACTB	FYN	PPP1R9B	ALAD	2210409H23RIK	ALOX12B	1810007A24RIK	HPN
APBB2	GAD1	PTN	ALAS2	4933402O15RIK	CASP14	1810009J06RIK	HSD3B1
APLP1	GAP43	PTPRZ1	BLVRA	4933425L11RIK	CST6	1810029G24RIK	INS1
CDH4	GFRA1	ROBO1	CPO	ACADS	EVPL	2200008D09RIK	INS2
CDK5	GNAQ	RTN1	EIF2AK1	ACLY	GJB5	2210010C04RIK	KLK5
CLDN11	GPM6B	S100B	ERAF	ADH7	KLF4	2810031L11RIK	LPIN1
CRMP1	GSS	SEMA3C	FECH	ALDO1	KLK7	ACOX2	NGFG
CRYM	HDAC5	SEMA4G	GATA1	ATF3	MEST	ADH1	PEMT
CXCL12	HES6	SEMA5A	HEBP1	ATP5B	OVOL1	AQP11	PNLIPRP1
DBN1	IGF1	SEMA5A	Hbb-b1	CS	SCEL	ARSA	PNLIPRP2
DCAMKL1	IGSF8	SERPINI1	HMBS	ENO2	SPRR1A	CASP9	PRKAG1
DCC	KIF5C	SIAT8B	PPOX	FDFT1	SPRR1B	CEL	PRSS2
DCX	LHX1	SPOCK2	SLC4A1	GAPD	MAFB	CLPS	PTER
DOK5	MAPK8IP3	STMN2	TAL1	GPD2	KRT1 10	CPA1	QDPR
DPYSL3	MOG	STMN3	UROD	GYG1	KRT1 12	CPB1	RNASE3
DPYSL4	MTAP2	TAGLN3	UROS	HK2	KRT1 13	CPN1	SREBF1
DRPLA	MYO1B	THBS1		LDH1	KRT1 17	CTRL	TGM2
EFNA5	NDN	TMOD2		LDH3	KRT1 19	CYB5	TRYGN16
EGFR	NES	TOP2B		NDUFA1	KRT1 2	CYP17A1	ZDHHC3
EPHA4	NEUROD1	TRIO		PFKFB1	KRT1 4	EBP	
EVL	NINAT	UNC5C		PGK1	KRT2 1	ELA1	
EVL	NRN1	WNT11		PGK2	KRT2 10	ELA3B	
FACL6	NRP			SEC14L2	KRT2 16	F10	
FEZ1	NTRK2			SOAT1	KRT2 18	FABP1	
FGFR1	OLFM1			SULT1A2	KRT2 19	FAH	
FMR2	PAFAH1B3			TPI	KRT2 6A	HABP2	
FOXG1	POU6F1			UQCRB	KRT2 6G	HAO3	

that, on the whole, transcription changes gradually from stage to stage.

PCA is a dimensional reduction algorithm which identifies common trends in data sets. When applied to this data set, PCA analysis identifies component 1 as a set of thousands of genes with gradually increasing expression across development and the converse group of thousands of gradually decreasing genes (Fig. 1B). This finding is in agreement with the hierarchical clustering results, showing that transcriptional changes during mouse embryonic development are largely gradual, with the major trends being slowly increasing or decreasing transcription across the time course.

Time course analysis of coordinated patterns of expression

While overall transcriptional change is gradual during development, many of the most interesting developmental events are likely to be heralded by more discrete characteristic alterations in transcription in groups of genes. We performed K-means non-hierarchical clustering on the 17,491 variable genes to identify clusters of genes with similar expression patterns. From this analysis, ten major clusters containing a total of 31 subclusters could be distinguished (Fig. 1C). Component genes of all clusters are presented in Supplemental Table 2. Cluster 1 represents 6493 genes whose expression declines throughout development, while Cluster 2 is composed of 4440 genes with increasing expression across the time course. Interestingly, these clusters overlap significantly with the genes identified by PCA as correlated (upward trending) or anti-correlated (downward trending) to principal component 1, showing that distinct analysis methods yield similar results.

The other defined clusters represent somewhat more complex gene expression patterns, with fewer component genes. As we shall discuss later, many of these patterns correlate with distinct developmental events. Cluster 3 is composed of 1928 genes with a characteristic jump in expression between e9.5 and e10.0 followed by stepwise increases to ~e12.0. Cluster 4 contains 709 genes which exhibit a dramatic rise in expression between e11.5 and e12.0. Cluster 5 (503 genes) exhibits a smooth, U-shaped arc of expression, gradually trending down from e9.0 to e13.0 then increasing to p1. Cluster 5 genes substantially overlap with genes of the second principal component in the PCA analysis. Cluster 6 (369 genes) is characterized by a significant drop in expression between e9.0 and e9.5 followed by slower decreases to ~e12.0. Cluster 7 (522 genes) genes have slowly increasing expression from e9.0 to ~e12.5, decreasing to ~e15.0, with a more rapid decline thereafter, while cluster 8 (306 genes) genes describe a steeper rise and fall with a peak at ~e15.5.

Clusters 9 and 10 are characterized by more dramatic spikes in expression. Transcription of cluster 9 (331 genes) genes is dramatically upregulated between e15.5 and e16.0 before dropping at e18.5. Genes in cluster 10 (379 genes) undergo a sharp induction of expression at e16.5 and e17.5 before declining at p1.

Timing of the activity of biological processes during embryonic development revealed through functional cluster analysis

To gain a more mechanistic understanding of the process the genes in each cluster represent, we used Fisher's Exact Test analysis to identify Gene Ontology (GO) functional categories which were significantly overrepresented in each of the defined K-means clusters and PCA components (Zeeberg et al., 2003). We will focus our discussion on highly significant biological process categories in these clusters which shed light on both broad and specific aspects of mammalian embryonic development and, subsequently, their temporal similarities with *Drosophila* embryonic development.

The sizes of cluster 1 (downward across the time course) and 2 (upward) suggested that there are major trends in gene expression which might reveal interesting insights into development as a whole. GO overrepresentation analysis of the downward trending cluster 1 demonstrates that a number of basic processes are significantly more prominent at early developmental times and wane later (Supplemental Table 1.1). These include "DNA and RNA metabolism," "cell cycle," "DNA replication, packaging, chromatin architecture," "nuclear organization and biogenesis," "transcription," "translation and translation initiation," "protein folding," and "pattern specification." The overrepresentation of these categories at earlier developmental times fits the postulated embryonic pattern of extremely rapid cell division and proliferation early, with substantial later withdrawal of many cell types from the cell cycle and consequent decreases in DNA replication and associated processes. Transcriptional processes are also more prominent early, with the expression of a large number of transcription factors fitting this profile. As noted, PCA analysis identified a group of genes anti-correlated to component 1 which substantially overlaps the genes of cluster 1 and exhibits a similar downtrend in expression. The overrepresented processes in this group are nearly identical to those of Cluster 1 (Supplemental Table 3.2) Thus, these two independent statistical analysis methods lead to the same conclusions.

Analysis of Cluster 2, the upward trending genes, highlights a contrasting set of processes which become more active as development progresses (Supplemental Table 1.2). Many of the most significant categories in this group are "energy pathways" with subcategories including "fatty acid metabolism," "electron transport," "TCA cycle," "oxidative phosphorylation," and others. "Cell adhesion" and related subcategories are also highly enriched, as are "muscle development" and "contraction," "defense response," and its subcategories involved in immune system development and function. Again, the group of genes correlated to principal component 1 (upward trending) in the PCA analysis substantially overlaps the genes of cluster 2 and yields nearly identical results in GO overabundance analyses (Supplemental Table 3.1).

Examination of categories in cluster 5, genes with a gentle "U"-shaped expression pattern with a minimum at ~e13.0, revealed a striking enrichment in genes dealing with sugar metabolism, including "glycolysis," "gluconeogenesis," and related categories (Figs. 1C, D, Supplemental Table 1.5). This

suggests that, while the major trend in energy metabolism is generally upward across this period of development, the complete picture is somewhat more complex and may reflect stage-specific conditions and requirements. In this case, genes from principal component 2 of PCA overlap with those of cluster 5, have a similar expression pattern, and yield similar GO category overabundance results (Supplemental Table 3.3).

Cluster expression patterns correlate with specific developmental events

As we analyzed the more complex K-means cluster patterns, it became clear that categories of genes associated with specific developmental processes were often characteristically overrepresented in particular clusters, contributing significantly to the cluster "shape." Among the most interesting is cluster 3, whose genes undergo a significant jump in transcription between e9.5 and e10, with subsequent increases to a plateau at ~e12.0. The most overrepresented biological processes in this group include "neurogenesis," "axonogenesis," "central nervous system development," and other categories related to neural differentiation and function (Figs. 1C, D, Supplemental Tables 1.3, 2.3). More than 80 genes from this cluster are known to be associated with neurogenesis, including well-known neural markers such as neurofilaments, *NeuroD1* (Kageyama et al., 1997), *NCAM*, *FoxG1*, *Hes6*, *Scg10* (*Stmn2*), *Robo1* (Sundaresan et al., 2004), *Ephrin A5* (Bolz et al., 2004), and *Eph A4* (Fig. 2A). The dramatic increase in transcription from this cluster of neural genes occurs at the exact time when the walls of the primitive brain and spinal cord begin to differentiate into three distinct layers at ~e10.0–e10.5 and shows remarkable coordination of transcription of a large neural specific set of genes (Kaufman and Bard, 1999). A significant number of relatively uncharacterized ESTs are included in this cluster; this population is undoubtedly highly enriched in novel molecules involved in neural differentiation. Interestingly, co-regulation of expression of a number of neuronal genes with a peak near mid-embryogenesis was also observed in the analysis of the *Drosophila* life cycle (Arbeitman et al., 2002) (Supplemental Fig. 1).

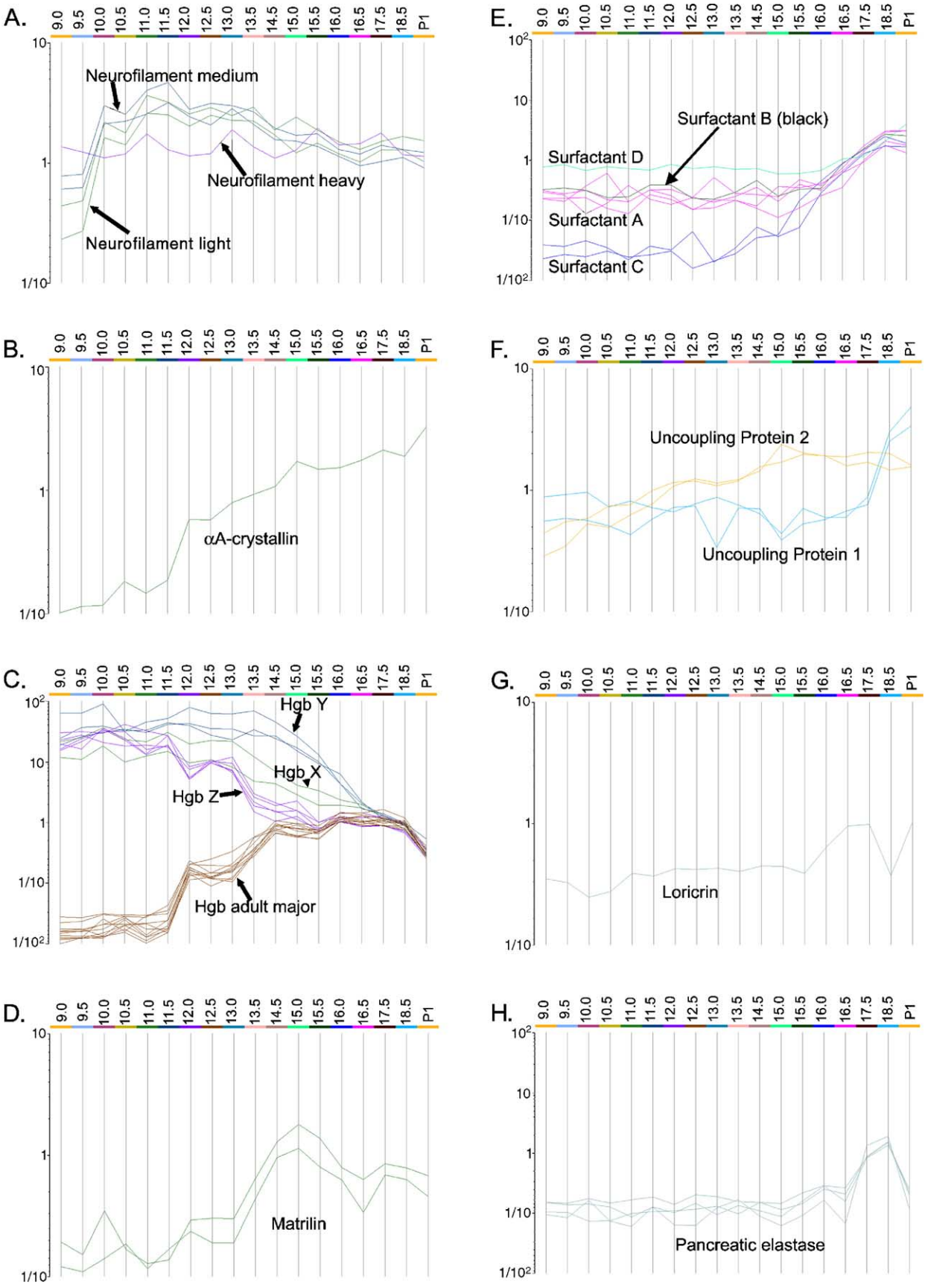
Another very interesting expression pattern is that of cluster 4, which is characterized by a dramatic induction of transcription between e11.5 and e12.0. The overrepresented biological processes in this cluster primarily involve heme biosynthesis and erythrocyte differentiation. Transcription of the major chain of beta adult hemoglobin clearly begins in earnest at e12.0 (see Fig. 2C), as it does for factors such as *Tal1*, *erythroid associated differentiation factor* (*ERAF*), *heme binding protein 1* (*HEBP1*), and several enzymes involved in hemoglobin biosynthesis (Koury et al., 2002) (Fig. 1D, Supplemental Table 1.4). This corresponds with the beginning of definitive hematopoiesis in the liver at e11.5–e12.0 when hematopoietic stem cells from the blood islands have colonized the septum transversum (Dzierzak and Medvinsky, 1995; Kaufman and Bard, 1999).

One of the most striking expression patterns identified by the K-means clustering is cluster 9, whose genes are transcribed at relatively low levels until ~e15.5–e16.0 but are then remarkably induced to a peak at e16.5–e17.5 before leveling off or dropping at e18.5 (Fig. 1C). Annotation analysis shows that "epidermal differentiation" and related biological processes are highly overrepresented in this cluster, with many genes involved in differentiation of the skin stratum corneum such as *Casp 14* (Rendl et al., 2002), *cystatin 6* (Zeeuwen et al., 2001), *envoplakin* (Leung et al., 2002), *periplakin*, *loricrin* (Hardman et al., 1998), *kallikrein 7* (Komatsu et al., 2003), *KLF4* (Jaubert et al., 2003; Segre, 2003), *MAFB* (Ogata et al., 2004), and *SPPR1* (Tesfaigzi and Carlson, 1999) as well as multiple keratin complex genes (Figs. 1D, 2G, expression pattern of loricrin). The timing of this induction is exactly coincident with terminal epidermal differentiation and permeability barrier formation in the mouse, as would be predicted. Interestingly, given the proposed role of Wnt pathway signaling in epidermis and hair follicle development, *Wnt4* and *Wnt7b*, both of which have been shown to be expressed in the skin, are also included in this cluster (Alonso and Fuchs, 2003a,b; Shu et al., 2002).

Another dramatic pattern reflecting closely aligned induction of related genes is exhibited by cluster 10, where genes are expressed at low levels before being induced dramatically at e17.5 and e18.5. In the overabundance analysis, the biological processes "digestion," "lipid metabolism," and "proteolysis" categories characterize a large number of these genes. Many of these genes are associated with exocrine pancreatic activities such as *pancreatic trypsin 2* and *4*, *pancreatic lipase-related proteins 1* and *2*, *carboxyl ester lipase*, *pancreatic colipase*, *pancreatic amylase 2*, *carboxypeptidase A1*, *chymotrypsin-like*, *pancreatic elastases 1* and *3b*, as well as endocrine pancreatic activity, exemplified by *insulin1*, *insulin2*, and *glucagon*, and pancreatic differentiation and development, exemplified by *regenerating islet-derived 1* (Fig. 2H). Again, induction of this suite of genes occurs just at the time of pancreatic terminal differentiation and the onset of exocrine and endocrine function at e17.5 and just in time to prepare the soon to be born animal to digest on its own (Johansson and Grapin-Botton, 2002; Kaufman and Bard, 1999).

Genes activated at specific developmental stages—validation of clustering results using SAM analysis

To help validate the findings from the cluster analyses, we used the Significance Analysis of Microarrays (SAM, FDR < 0.10) algorithm to look for sets of genes whose expression abruptly changes in a statistically significant manner between two developmental timepoints (Supplemental Table 4). The results of this analysis confirmed many of the findings from K-means clustering. For example, K-means clustering identified cluster 3 as heavily weighted toward neural differentiation genes upregulated at e9.5–e10.0. At this same timepoint, SAM analysis identified 60 upregulated transcripts, of which at least 24 are involved in neural differentiation. Between e11.0/e11.5 and e12.0/e12.5, SAM identifies 384 upregulated genes, including the adult hemoglobins and erythrocyte developmental proteins present in cluster 4. There is also a substantial


A.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10
1
1/10
Neurofilament medium
Neurofilament heavy
Neurofilament light
B.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10
1
1/10
αA-crystallin
C.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10²
10
1
1/10
1/10²
Hgb Y
Hgb X
Hgb Z
Hgb adult major
D.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10
1
1/10
Matrilin
E.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10²
10
1
1/10
1/10²
Surfactant B (black)
Surfactant D
Surfactant A
Surfactant C
F.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10
1
1/10
Uncoupling Protein 2
Uncoupling Protein 1
G.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10
1
1/10
Loricrin
H.
9.0 9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5 14.5 15.0 15.5 16.0 16.5 17.5 18.5 P1
10²
10
1
1/10
1/10²
Pancreatic elastase

overlap between cluster 9 and the upregulated SAM genes between e15.0/e15.5 and e16.0/e16.5, with many of the same keratinocyte maturation genes present in both. SAM analysis also identifies the dramatic increase in transcription of digestive enzymes and hormones between e16.0/e16.5 and e17.5/e18.5 that forms the basis of K-means cluster 10. Again, these findings show that independent analysis approaches can be used to identify the same developmental events, supporting the strength of the conclusions.

Evolutionary conservation of embryonic developmental gene class usage patterns

Transcriptional studies of development in other species have yielded interesting insights about not only single genes, but also *classes* of genes and the timing of their utilization during embryogenesis. Of particular interest to us was the elegant analysis of the *Drosophila* life cycle (Arbeitman et al., 2002) in which major temporal biases in the expression of classes of genes during embryogenesis were noted, revealing a transcriptional superstructure of development in the fly. Given the high degree of conservation of genes, gene functions, and pathways between *Drosophila* and mouse, we were interested in systematically analyzing whether this transcriptional superstructure was also conserved and in how such conservation might provide insights into the evolution of developmental mechanisms across the 550 million year gulf which separates the two organisms. To do this, we compared gene class expression patterns during mouse embryogenesis with each of the six GO classes which exhibited clear patterns during *Drosophila* embryonic development (Supplemental Fig. 2).

The analysis of *Drosophila* embryonic development identified a number of coordinately regulated classes of genes representing biological processes which are highly active during earlier stages of embryogenesis, but which decline as the embryo nears the transition to the larval state. Given that both early *Drosophila* and mammalian embryos are composed of relatively few relatively undifferentiated cells that must subsequently produce large numbers of differentiated cells from which to create specific structures, we examined whether classes of such as “cell cycle” and “transcription factor” likely involved in these basic processes had similar temporal patterns of utilization. During *Drosophila* embryogenesis, transcription of cell cycle genes as a class is very high at the earliest timepoints but declines steadily until the larval transition. We observe a similar pattern of expression in the mouse embryo, with G1 and G2 cyclins and their CDKs highly expressed earlier in embryogenesis but declining later (Fig. 3A, Supplemental Table 1.1). Many other genes involved in cell cycle and chromosome replication functions such as CDC homologs, mini-chromosome maintenance (MCM) deficient genes, and structural maintenance of chromosome (SMC) genes are also seen to be more highly expressed early in development (Fig. 3B). Several exceptions included in the analysis serve as controls. For example, “cyclins” G and G2 have a very distinct expression pattern, peaking in late development. Why is their expression different from the other cyclins? *Cyclin G* was assigned its name due to homology with other cyclins but does not function as a cyclical regulator of cell cycle progression but rather as a target recruited upon p53 activation, which would be expected to be expressed more robustly later in development (Okamoto et al., 2002). *CDC25b*, unlike other CDC homologs, has a later peak of expression which likely reflects its role as a steroid hormone receptor coactivator (Chua et al., 2004; Ngan et al., 2003). The *G0/G1 switch gene 2*, also with a late peak, is expressed primarily in maturing lymphocytes and not other cell types (Russell and Forsdyke, 1991).

As expected, the expression of cyclin/CDK inhibitors such as *p15*, *p18*, *p19*, *p21*, and *p57* is the converse of that of the cyclins they inhibit, with quite low expression early which increases as development proceeds (Fig. 3C). Again, an interesting internal control for these observations is included in the form of *p27*, a G1 cyclin/CDK inhibitor which is known to be primarily post-transcriptionally regulated and which shows little variation in transcript levels across the time course (Hengst and Reed, 1996).

Two other major classes of genes were also noted to be clearly expressed preferentially during early *Drosophila* embryogenesis with tapering expression later: transcription factors and chaperone proteins (Supplemental Fig. 2). These patterns are retained during murine embryogenesis as well, where transcription factors and protein folding genes are also preferentially expressed early in development (Supplemental Table 1.1). Again, upon consideration of the tasks required of the early embryo of both organisms, which include the extremely rapid production of cell constituent proteins in the service of rapid cellularization and proliferation, it is not surprising that there has been strong selective pressure to retain these patterns as well.

Another clear class transcriptional trend that was noted in the *Drosophila* study is the converse early very *low* class-wide transcription of cell adhesion molecules, which are then steadily and dramatically induced as the embryo approaches hatching (Supplemental Fig. 2). This pattern fits with known features of *Drosophila* development. At early stages, the embryo is a syncytium protected by the egg shell, and there

Fig. 2. The data set recapitulates the developmental regulation of gene expression at multiple timepoints during embryonic development. Where possible, data from duplicate array elements are provided to demonstrate the degree of concordance between elements representing the same gene. Data presented as expression ratio on a $\log_{10}$ scale. (A) Onset of expression of *neurofilament medium* and *light chains* at ~e9.5–e10.0. (B) *αA-crystallin* expression begins in the lens cup at ~11.5. (C) The onset of definitive hematopoiesis at ~e12.0 is evident from the expression pattern of adult hemoglobin. Transcription of the *embryonic X*, *Y*, and *Z hemoglobins* declines in characteristic patterns as development progresses. (D) Onset of *matrilin* expression coincides with chondrocyte maturation and exit from the proliferative phase at ~e13.0–e13.5. (E) Pattern of lung surfactant expression just prior to birth. (F) *Uncoupling protein 1* is required for non-shivering thermogenesis and is greatly induced at e18.5. (G) A dramatic peak in transcription of the keratinocyte gene *loricrin* at ~e16.5–e17.5 is representative of the expression pattern of many epidermal specific genes in cluster 9. (H) Expression pattern of *pancreatic elastase* in the developing embryo. Multiple pancreatic digestive enzyme and hormone genes in cluster 10 exhibit a similar pattern.

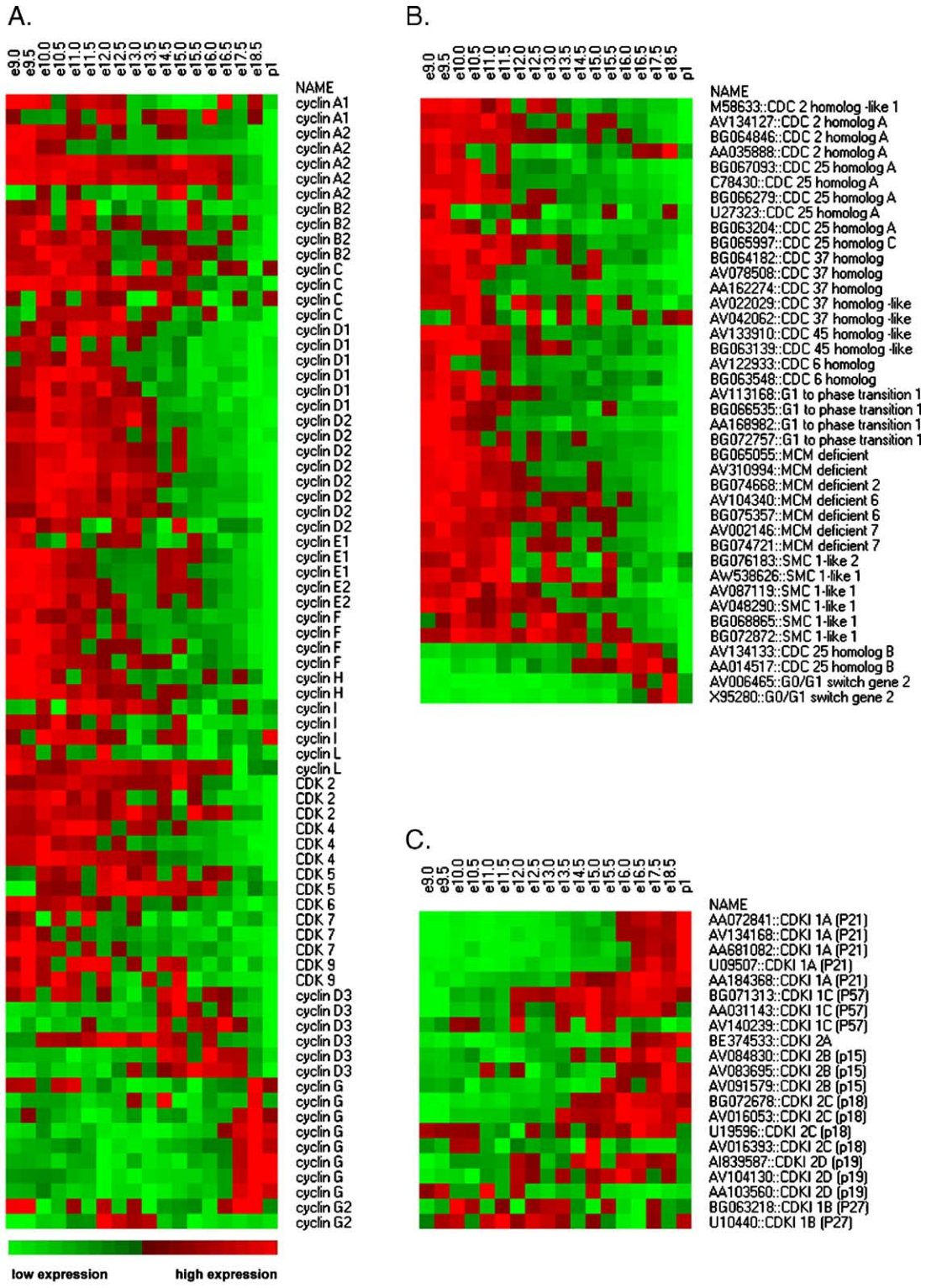


Fig. 3. Expression of cell cycle genes during mouse embryonic development. (A) Heatmap of cyclin expression across the time course. Cyclins are transcribed at higher levels earlier in embryogenesis, with declining expression as the embryo matures. The cyclin G genes are not involved in cell cycle control but are targets of p53 activation. (B) Other genes involved in DNA and cellular replication are more highly expressed during earlier embryogenesis. (C) The expression pattern of the cyclin/CDK inhibitors is opposite to that of the cyclins they inhibit, consistent with a general slowing of cell cycle progression during later development.

is little need for an extracellular structural framework. Later, as the time nears for the larva to venture into the world, cells must remain oriented within compartmental boundaries, and the organism must become more robust to handle gravitational and mechanical stresses, functions which require the presence of such a framework.

The expression of cell adhesion molecules in the mouse embryo exhibits the same temporal pattern. Again, while the

mouse embryo at the earliest stages assayed here has differentiated into three cell layers, there is little need early on for a very rigid extracellular framework since it is supported and cushioned by extraembryonic structures and fluid. Indeed, such a rigid framework would be a hindrance to the many cellular migration processes still occurring during this period of development. However, like *Drosophila*, as the mouse embryo matures, it must transform from a loose aggregation of cells into a much more sturdy assemblage in which there are tissue and compartmental boundaries and which can interact with the external environment and move without pulling itself apart and thus must produce the collagens, fibronectins, cadherins, and other cell adhesion proteins necessary for this function. Beyond these genes involved simply in cell adhesion, the class of "structural" genes which includes many actins and actin modifiers, tubulins, myosins, and other muscle contractile gene, etc., are also expressed at very low levels during earlier stages of embryogenesis in both *Drosophila* and the mouse, with a dramatic rise as hatching or birth approach.

There is interesting evidence that the two types of embryo handle another "birth"-related stress in a similar way. One of the most dramatic class expression trends that is seen in both *Drosophila* and mouse embryogenesis is a concerted upregulation of metabolic genes just prior to the end of embryogenesis (Supplemental Fig. 2). In the case of *Drosophila*, many of the genes that are activated are involved in mitochondrial oxidative phosphorylation and ATP synthesis, and an extraordinarily similar profile is seen in the murine embryo, where orthologous genes from nearly all complexes of oxidative phosphorylation are upregulated, as is fatty acid metabolism (Supplemental Fig. 3). What is the value of this conserved mechanism to both organisms? Each must prepare beforehand to expend significantly greater amounts of energy after birth, the larva for muscle activity, mobility, and acquisition of food, and the mouse pup for breathing, muscle activity, gastrointestinal activity, etc., as it matures. Interestingly, as *Drosophila* enters the pupal state, its metabolic gene expression reverts to the early embryonic pattern, again low early in pupation and increasing as eclosion approaches, with the onset of this metabolically less active state anticipated by downregulation of the whole suite of genes late in larval life, suggesting a coordinated transcriptional regulation mechanism. There is evidence for similar coordinated rapid reactivation of fetal metabolic transcriptional patterns in the stressed mammalian myocardium (Ashrafian et al., 2003; Stanley and Chandler, 2002; Wagner et al., 2004; Zhang, 2002), where coordinated transcriptional downregulation of dozens of oxidative phosphorylation and fatty acid oxidation genes can take place as rapidly 4–12 h after the insult (R. Wagner, data not shown).

Developmental regulation of muscle differentiation genes—conservation of a discrete system

Muscle differentiation is an evolutionarily ancient process which has retained many mechanisms across evolutionary time and which has been examined in some detail in the microarray experiments describing *Drosophila* development (Arbeitman et al., 2002). In both *Drosophila* and mammals, muscle develops in segmental blocks along the longitudinal axis of the embryo. In *Drosophila* embryogenesis, this differentiation takes place at one time in a coordinated manner (Paululat et al., 1999). In mammals, the paraxial mesoderm from which the segmental somites develop differentiates in a rostral to caudal temporal progression. Thus, the predicted developmental pattern of expression for muscle differentiation genes (or characteristic genes for any segmentally expressed somite derive tissue) in the whole mammalian embryo would be a slow onset of expression as the first early somites differentiate, with increasing levels as more somites follow and the muscle compartment expands. The K-means analysis identified two subclusters within the uptrending cluster 2 (subclusters 2a and 2b, Supplemental Tables 1.2a and 1.2b) which are especially enriched for muscle development and structural genes; the general expression pattern of these clusters is one of gradually increasing expression across the time course, with an upward inflection beginning at ~e14.0 with a steeper upward slope thereafter corresponding with the substantial expansion of muscle mass which occurs in the embryo at this time. Fig. 4A shows a heatmap of muscle developmental and structural genes, demonstrating their increase in transcription in late embryonic life. A heatmap of several specific genes utilized at specific stages of muscle maturation shows this developmental progression more clearly (Fig. 4B). *MyoD* is a bHLH muscle regulatory factor (MRF) transiently expressed in the second wave of somitic myogenesis beginning at e9.5 (Smith et al., 1994). *Myocyte enhancer factor 2C* (*Mef2c*) is a secondary transcription factor induced by the MRFs, and *myogenic factor 6* is the main MRF expressed in myotubes and adult muscle (Yoon et al., 1997). We also examined expression of several contractile component genes including skeletal slow troponin T1, which is known to be expressed as early as e10.5–e12.0 on (Krishan et al., 2000), *embryonic skeletal myosin heavy polypeptide 3*, *perinatal skeletal muscle myosin heavy polypeptide 8*, *fast skeletal myosin light chain*, and *skeletal muscle myosin heavy polypeptide 4*. *MyoD1* transcription is enhanced first at e9.5 followed by a slow increase in transcription of *MEF2C* and *troponin T1*. As muscle matures, there is increasing expression of *Myf6*, as well as the onset of expression of structural muscle markers in the appropriate developmental order.

The data set recapitulates known temporal regulation of gene expression and highlights differentiation events

We were interested in asking the question of how closely our microarray results corresponded with known expression patterns, both as a validation of the data and as a demonstration that an interesting expression pattern may be a useful guide for suggesting novel genes for further study. Inspection of the data set for known developmentally regulated gene expression patterns shows that our results are consistent with reported expression data from the literature for a large number of genes representing multiple different tissues. For example, neurofilament proteins are highly specific markers of neuronal differentiation whose tentative first expression is seen in a

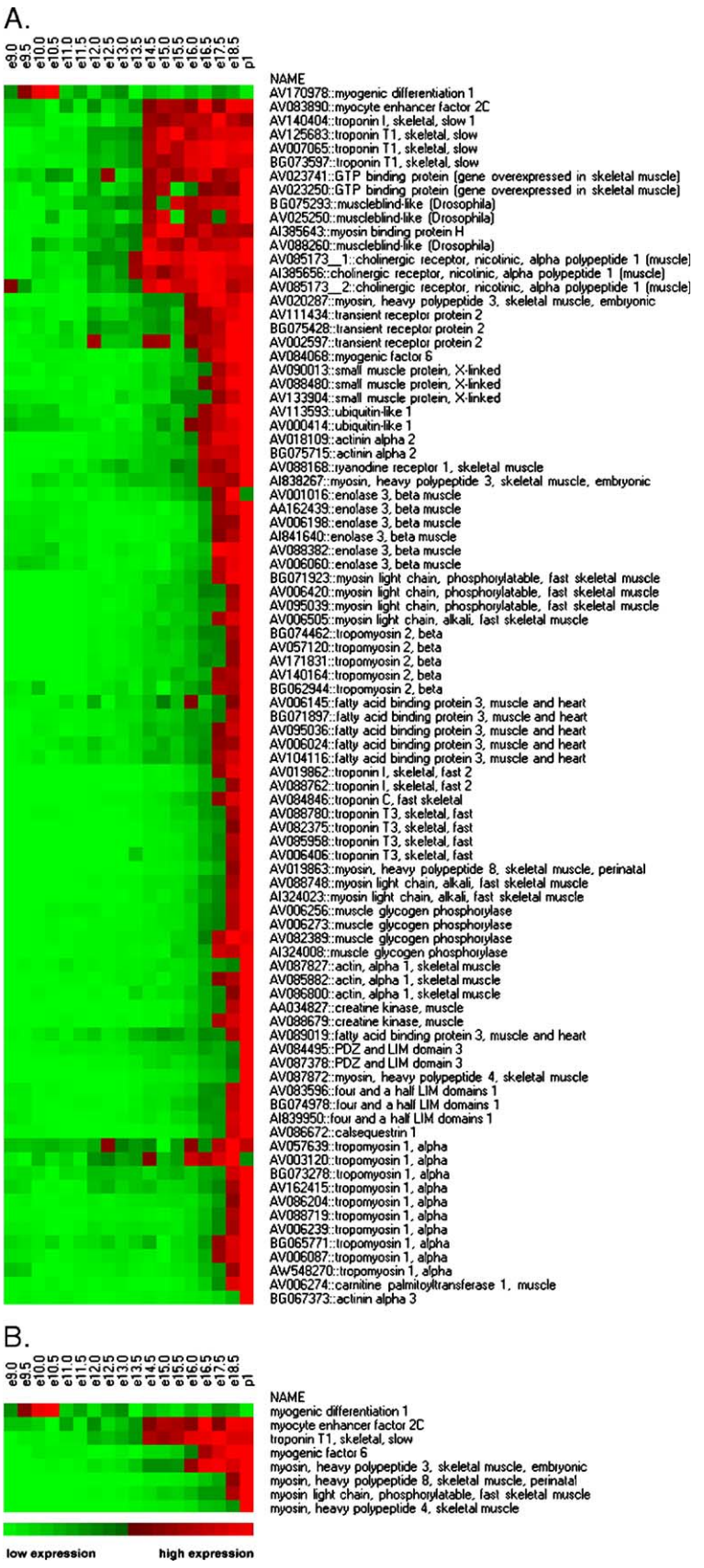


Fig. 4. Muscle differentiation in the developing embryo. (A) Heatmap of expression of skeletal muscle markers in the developing embryo. *MyoD1* is the earliest assayed marker and has a peak in expression at ~e10.5. Downstream markers of differentiation appear later in the expected progression. (B) Condensed heatmap more clearly demonstrates the sequential expression of muscle differentiation markers in the expected temporal progression.

mere handful of cells in the central nervous system, neural tube, and peripheral ganglia at e9.5, but whose expression is robust and much more widespread by e10.5, as demonstrated elegantly in several classic experiments (Cochard and Paulin, 1984; Edwards et al., 1989; Julien et al., 1986). Our array data agree with these known patterns, showing dramatic induction of transcription of *neurofilament light* (green lines) and *medium* (blue) polypeptides, as expected, but not of the *heavy* (purple) polypeptide, whose expression is known to increase substantially only after birth (Fig. 2A).

The vertebrate lens expresses a number of crystallins, with *αA-crystallin* being the predominant protein. These crystallins function as lens-specific structural and light refracting proteins, with additional chaperonin activities as well. Expression of *αA-crystallin* has been shown to begin in a few cells in the lens cup at e10.5 with a significant increase in expression in the lens vesicle by e11.5–e12.5 (Robinson and Overbeek, 1996), a pattern that is clearly reflected in the array data (Fig. 2B).

One of the best known mammalian developmental expression patterns is that of the fetal and adult globin genes. During early embryonic development, fetal hemoglobin isoforms with high oxygen affinity are expressed to optimize maternal-to-fetal oxygen exchange, but, as the embryo matures, fetal isoforms are gradually replaced by adult isoforms following the onset of definitive hematopoiesis at ~e12.0 (Fig. 2C). The array data clearly demonstrate the onset of transcription of *beta adult hemoglobin* (brown lines) message at e12.0, as well as the slow decline of messages for *Hgb Y, beta-like embryonic* (green), *hemoglobin X, alpha-like embryonic* (blue), and *Hgb Z, beta-like embryonic* (purple) from e11.5 on as the fetal chains are replaced by the adult isoform.

During endochondral bone formation, chondrocytes progress through a differentiation program before being replaced by bone. Chondrocyte maturation and exit from the proliferative phase are marked by the onset of expression of *matrilin-1* (*cartilage matrix protein*) at ~e13.0–e13.5, with strong expression at e15.5 which declines somewhat by e17.5 as chondrocytes are replaced by ossified bone (Aszodi et al., 1994). Again, the expression profile of *matrilin-1* derived from our array data matches this known pattern closely (Fig. 2D).

A major complication of premature birth is respiratory distress syndrome caused by a lack of surfactant protein in the lungs. In the mouse embryo, differentiation of type II alveolar cells and the onset of expression of surfactant proteins is known to occur beginning at ~e15.5–e16.0, and the array expression data are consistent with this, showing substantial induction of *surfactant associated protein C* (Fig. 2E, blue lines) at e15.5. Induction of *surfactant associated proteins A* (pink), *B* (black), and *D* (green) occurs slightly later at e16.0, a progression which has been well documented in previous studies (Mendelson, 2000; Whitsett and Weaver, 2002; Yang et al., 2002; Zhou et al., 1996).

Differentiation of the serous cells of the salivary glands is known to begin at around the time of birth and is necessary for the production of secretions required to wet the oral mucosa and allow feeding. Expression of two markers of serous cell differentiation, *parotid secretory protein* (*Psp*) and *demilune cell and parotid protein* (*Dspp*), has been shown to increase dramatically in the early postnatal period (Poulsen et al., 1986). Our array data show that this expression actually begins slightly earlier, with a dramatic rise in transcription in the embryo beginning at e17.5 and continuing through p1, consistent with both the timing of serous cell differentiation and the need for production of these proteins (Supplemental Fig. 4).

The mitochondrial uncoupling protein *UCP1* is expressed in brown adipose tissue and is the major mediator of nonshivering thermogenesis. Its function is to divert proton motive force during oxidative phosphorylation by catalyzing a proton leak across the mitochondrial inner membrane, causing a portion of ATP synthesis to be bypassed and the energy to be released as heat. In accordance with the known expression of *UCP1*, our array data indicate that *UCP1* (Fig. 2F, blue lines) transcription is upregulated by over 10-fold at e18.5, just preceding birth, at just the time when the embryo must prepare to maintain body temperature on its own (Echtay et al., 2002; Jacobsson et al., 1985; Nedergaard et al., 2001; Ricquier and Bouillaud, 2000; Villarroya et al., 2001). In contrast, the transcription of the related gene *UCP2* (yellow), which does not significantly contribute to thermogenesis, does not change significantly at this time, appearing to fit with the slowly increasing general mitochondrial expression pattern.

The genes which were examined for validation are expressed at different times during embryonic development and at different levels, represent a broad range of different functions, and accurately reflect transcriptional changes observed in previous studies across a range of expression levels. This indicates that the expression data derived from these experiments are accurate and complete enough to serve as a valuable reference for the exploration of gene expression during this portion of embryonic development.

Genes from important gene families are expressed in novel patterns—hypothesis generation from the data set

Because of our interest in molecules involved in developmental patterning, we examined the data for distinctive expression patterns of genes from important pathways and will present examples of several of the more striking of these patterns. Given the complexity of the data set, these represent only a few of the most interesting patterns but are illustrative of the value of these data in hypothesis generation.

Zinc finger proteins are transcription factors which regulate a broad range of processes (Ladomery and Dellaire, 2002). While many zinc finger protein genes have been identified by sequence similarity, functional data exist for very few. Given the multiplicity of genes, such information would potentially be of great utility in determining which are worthy of further study for their involvement in specific developmental processes. Fig. 5A shows the expression of the annotated zinc finger genes on the array across the developmental time course. Note that most of the genes are expressed more highly at earlier times, consistent with the K-means cluster and GO annotation analyses which identified transcription factors as being more highly expressed earlier in development (cluster 1).

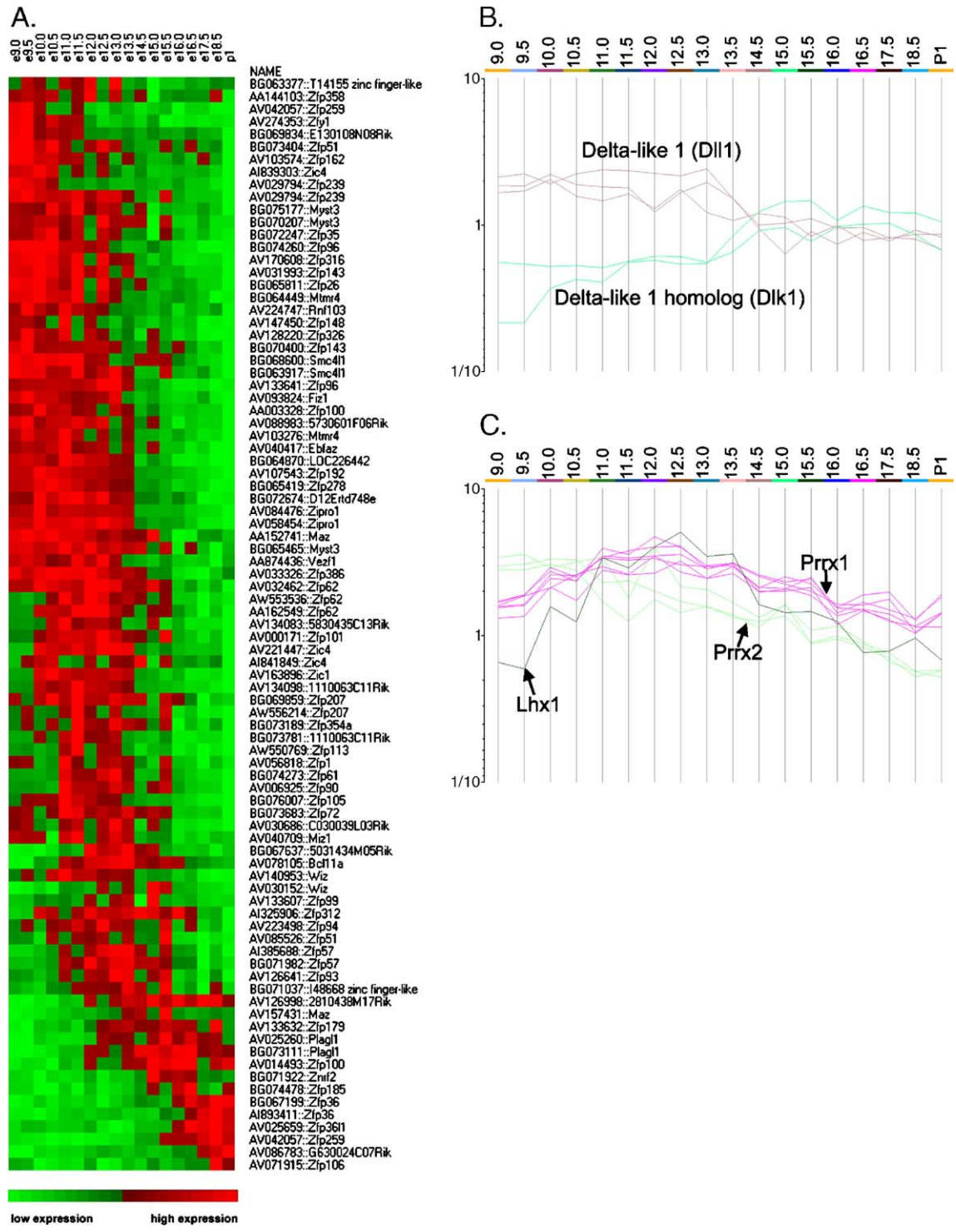


Fig. 5. Expression of zinc finger genes and developmental regulators during mouse embryonic development. (A) Heatmap of zinc finger gene expression across the time course. Most of these transcription factors are predominantly expressed earlier in embryogenesis, but a subset have later peaks which may implicate them in specific developmental events. (B) Expression patterns of *Delta-like 1* and *Delta-like 1 homolog*. (C) Expression patterns of three homeobox genes, *Prrx1*, *Prrx2*, and *Lhx1*, show distinct but different peaks in expression.

There are, however, subsets of zinc finger genes having peaks of expression at later times during development, including some with peaks in the perinatal period. The overlap of these peaks of expression with specific known developmental events raises the possibility that a specific zinc finger protein may participate in some aspect of a specific developmental event. Similar analyses of other interesting gene families will also suggest candidates for roles in specific processes, likely leading to the discovery of interactions that otherwise might remain hidden.

Another interesting example is that of the expression of the two delta gene homologs on the array (Fig. 5B). Similarly

named, the two genes are *Delta-like 1* (*Dll1*) and *Delta-like 1 homolog* (*Dlk1*). *Dll1* is expressed at relatively high levels until ~e13.0 before dropping steadily until e14.5. *Dlk1*, conversely, undergoes a significant increase in expression between e13.0 and e14.5. Some evidence suggests that *Dll1* plays a role in regulating the growth of myeloid and lymphoid cells, (Ohishi et al., 2002; Tohda et al., 2003) while *Dlk1* may be involved in chromaffin cell differentiation, which occurs from ~e13.0 to e16.5, but the major roles of these genes in development remain to be discovered (Huber et al., 2002; Van Limpt et al., 2003).

Homeobox genes play important roles in developmental patterning. The *paired-related homeobox 1* and *2* (*Prrx1*, *2*) genes are necessary for normal limb and craniofacial development from e9.5 on and have similar but not identical developmental expression profiles with peaks of expression at e12.0 and e9.5, respectively (Fig. 5B) (Lu et al., 1999a,b). Another homeobox gene, *Lim homeobox 1* (*Lhx1*), which is required for forebrain and head development, is dramatically transcriptionally upregulated at e10.0 and peaks at e12.5 before declining steadily (Moreno et al., 2004; Shawlot and Behringer, 1995).

Public database

The data presented here have great potential for use by molecular and developmental biologists studying a broad range of problems. In addition to making raw data available for further bioinformatics analysis, we have constructed a web-based database that provides easy access to the data for biologists (http://www.mousedevelopment.org). A continuously updated annotated expression graph for any gene on the array can be accessed by accession number, gene name or symbol, LocusLink ID, gene description, or sequence BLAST (Fig. 6).

Conclusions

Embryonic development is a very complex process, and any substantial deviation from the normal progression is incompatible with life. This suggests that the most basic processes and patterns are likely to have been conserved over long evolutionary distances. While embryonic development differs in myriad specifics between bilateral metameric animals, we do, in fact, know from hundreds of years of

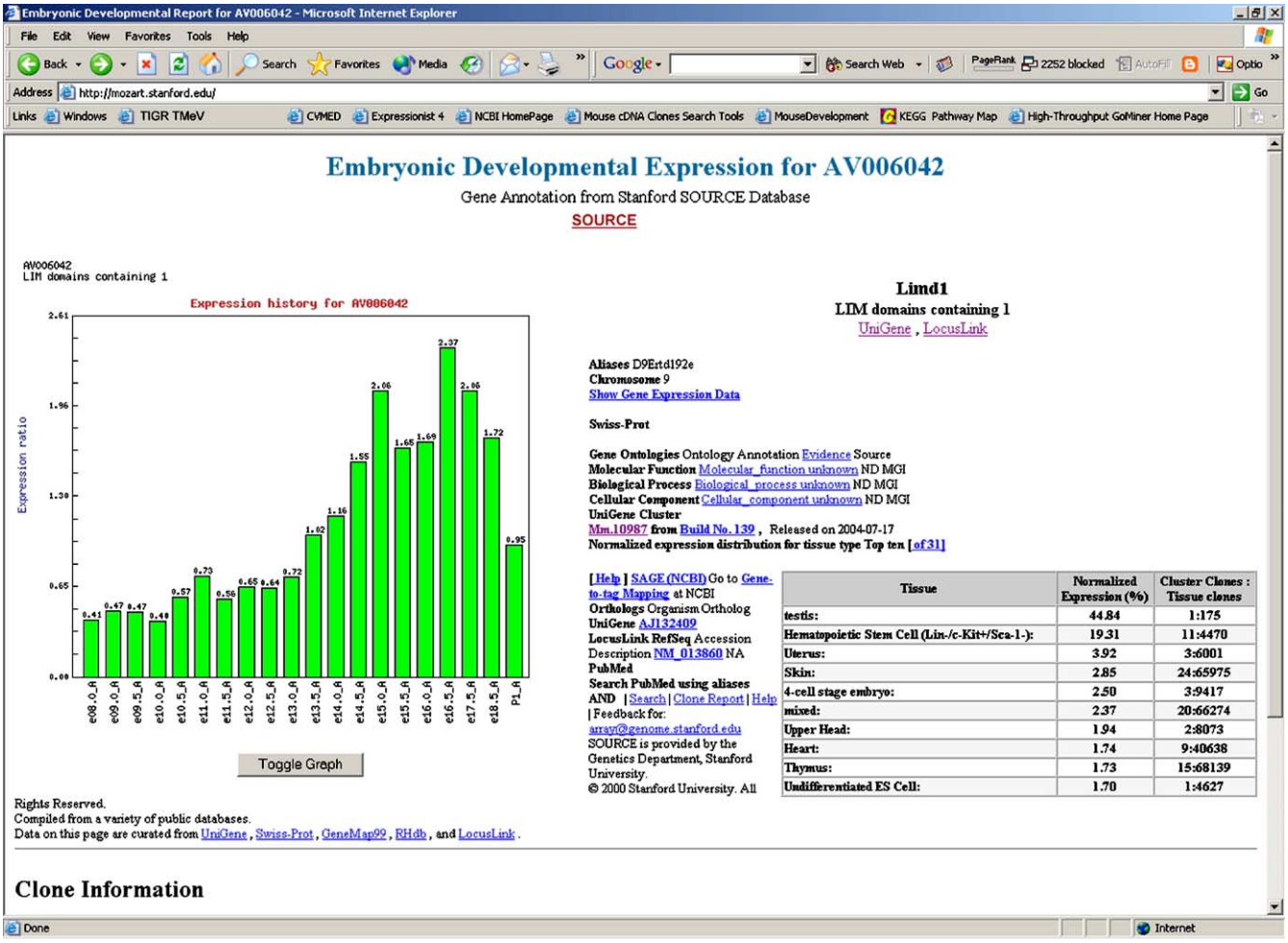


Fig. 6. Expression database. Screenshot of the results page from our publicly available database located at http://www.MouseDevelopment.org. The developmental expression profile of any gene on the array can be accessed by Unigene name or ID, LocusLink ID, or BLAST, with continuously updated annotation from the Stanford SOURCE database. Shown is the profile for the representative gene *Limd1*. Links to other major databases facilitate investigation of genes of interest.

biological study that, remarkably, major basic components of development have been retained since the time of an ur-bilateran common ancestor. Each arises from a single egg cell, each forms a blastula, then undergoes a form of gastrulation with the specification of three germ layers. In each, cells from these germ layers migrate and further differentiate, often in recognizably metameric patterns. Many examples have taught us that specific conserved orthologous genes are involved in mediating specific aspects of each of these processes in homologous tissues from *Drosophila* (and beyond) to mammals. Does this conservation extend beyond morphogenetic patterns and transcriptional control of isolated single genes to the level of conserved transcriptional regulation of whole related classes of genes involved in basic processes? Until now, with the advent of whole genome sequencing and microarray transcriptional analyses, we have been unable to answer this question. When we compare patterns of gene class usage across this period of mouse embryogenesis with *Drosophila* embryogenesis (Arbeitman et al., 2002), we *do* note striking similarities in the timing of the usage of whole classes of genes (Supplemental Fig. 2). Genes involved in a number of processes such as cell cycling, DNA synthesis and packaging, transcription, and protein folding are expressed preferentially during earlier stages of embryogenesis in both organisms, while those involved in cell adhesion, energy metabolism, mitochondria, and structural processes are preferentially utilized near the end of embryonic development. These similarities suggest that, as with the morphological aspects of embryonic development, the transcriptional superstructure of embryonic development arose once and has been conserved over roughly 550 million years of evolution via selective pressure to retain an effective basic design, with developmental elaborations that allow increased complexity built upon this solid foundation. Alternatively, convergent evolution of similar processes in response to common developmental requirements could give a similar result. While we support the hypothesis that modern developmental transcriptional programs represent elaborations on a basic ur-bilateran developmental transcriptional core, the final resolution of this intriguing question awaits further investigation and analysis, including in depth comparisons of transcriptional regulatory sequences and microRNAs (Grün et al., 2005; Krek et al., 2005) among the species which could reveal details of specific mechanisms of coordinate regulation.

Using several complementary statistical methods which yielded very similar conclusions, we have identified groups of genes of related function which are coordinately regulated to achieve differentiation and maturation events vital to the development of the embryo (Fig. 7). Exploration of the data have highlighted the coordinated regulation of many genes involved in distinct developmental events such as the onset of neuronal differentiation, definitive hematopoiesis, muscle differentiation and maturation, epidermal differentiation, and terminal differentiation events in the pancreas and digestive system. Given these distinct patterns, it is likely that study of other genes and ESTs with quite similar expression patterns will lead to the discovery of new players in these developmental processes.

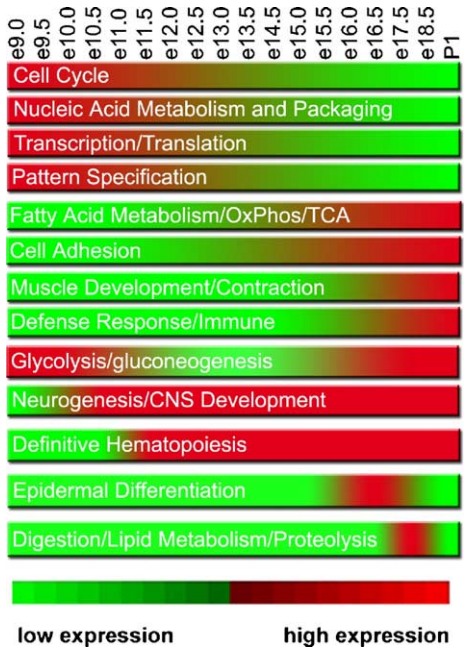


Fig. 7. Overview of activity of biological processes during mouse embryonic development. A number of processes including cell cycle, nucleic acid metabolism, transcription and translation, and pattern specification are more active at earlier stages in embryogenesis, while energy metabolism and differentiation processes predominate later.

During evolution, one of the mechanisms that has allowed vertebrates to gain substantially more complexity without similarly large increases in total numbers of genes is the reutilization of genes for diverse functions in different tissues. It is interesting that, despite this tendency, we are still able to identify so many distinct events from a transcriptional analysis in the whole mammalian embryo. However, gene reutilization likely obscures clues about other distinct events occurring in one tissue due to averaging of expression across the whole embryo and may particularly obscure tissue-specific activation of components of commonly used pathways such as Wnt, TGF, BMP, and others.

A second characteristic of mammalian development also likely interferes with the appearance of dramatic off–on switching of genes. In vertebrates, the paraxial mesoderm differentiates in a rostral to caudal temporal progression, with early somite derivatives undergoing differentiation at the same time when more caudal tissue is still undifferentiated, unsegmented mesoderm. This developmental organization leads to the appearance of a slow graded increase in transcription of genes from somite derived tissues, as exemplified by the expression of muscle-specific genes. In contrast, muscle differentiation in each segmental unit of the fly embryo takes place nearly simultaneously, resulting in a more distinct onset of transcription (Paululat et al., 1999).

We have assembled a cDNA microarray optimized for the analysis of mouse development and used it to systematically assess gene expression in mouse embryos from e8.0 to

postnatal day 1, during which many of the most important events of organogenesis occur. We limited our analysis to timepoints of e8.0 or later at which sufficient RNA could be prepared to directly label cDNA for array hybridization to avoid potential problems associated with linear amplification procedures. Previous extensive validation of array results using the same arrays and methods has been performed using quantitative real-time reverse-transcription polymerase chain reaction, and the array results have been confirmed for essentially all genes tested (Tabibiazar et al., 2003, 2005a,b; Wagner et al., 2004). Additionally, the array results are validated by their tight correspondence with a large number of previously documented developmental expression patterns, suggesting that the microarray data are accurate and reliable. Therefore, based on the use of the same arrays, methods, and stringent statistical parameters, further validation of these microarray results by qRT-PCR is not required (Hamatani et al., 2004; Ko et al., 2000; Wang et al., 2004). This study has produced gene expression data that are comprehensive and quantitative and complements elegant analyses of preimplantation embryonic development (Hamatani et al., 2004; Wang et al., 2004; Zeng et al., 2004) as well as analyses of embryonic development in other model organisms (Arbeitman et al., 2002; Birnbaum et al., 2003; Furlong et al., 2001; Hill et al., 2000; Jiang et al., 2001; Kim et al., 2001). The findings are the result of careful experimental design and stringent statistical analyses using multiple complementary approaches and are consistent between multiple replicates, with known global expression patterns, and with the behavior of known genes. We have touched on only a few of the intriguing findings present in this rich compendium, and further analysis will undoubtedly reveal additional insights. As a whole, the data provide a baseline for developmental biologists studying gene regulatory networks and systems biology. At the individual gene level, the data are an important and reliable foundation for the analysis of many aspects of mammalian embryonic and organ development and provide a starting point for the study of disease models, transgenic, and knockout mice in the continuing quest to understand the mechanisms of development.

Acknowledgments

R.A.W. was supported by grant HL076445 from the NHLBI. R.T. and T.Q. were supported by the Reynolds Center for Cardiovascular Research. The Stanford Mouse Transcriptome Microarray was developed in collaboration with Dr. Mike Fero and the Stanford Functional Genomics Facility. We sincerely thank Dr. Daniel J. Estes for assistance creating the online database, and Dr. Barry Zeeberg and Dr. Haiying Qin for assistance with High-Throughput GOMiner.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ydbio.2005.09.036.

References

Agrawal, D., Chen, T., Irby, R., Quackenbush, J., Chambers, A.F., Szabo, M., Cantor, A., Coppola, D., Yeatman, T.J., 2002. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. J. Natl. Cancer Inst. 94, 513–521.

Alonso, L., Fuchs, E., 2003a. Stem cells in the skin: waste not, Wnt not. Genes Dev. 17, 1189–1200.

Alonso, L., Fuchs, E., 2003b. Stem cells of the skin epithelium. Proc. Natl. Acad. Sci. U. S. A. 100 (Suppl 1), 11830–11835.

Arbeitman, M.N., Furlong, E.E., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., White, K.P., 2002. Gene expression during the life cycle of *Drosophila melanogaster*. Science 297, 2270–2275.

Ashley, E.A., Spin, J.M., Watt, C. (2004). Heatmap Builder 1.0. http://quertermous.stanford.edu/heatmap.htm.

Ashrafian, H., Redwood, C., Blair, E., Watkins, H., 2003. Hypertrophic cardiomyopathy: a paradigm for myocardial energy depletion. Trends Genet. 19, 263–268.

Aszodi, A., Modis, L., Paldi, A., Rencendorj, A., Kiss, I., Bosze, Z., 1994. The zonal expression of chicken cartilage matrix protein gene in the developing skeleton of transgenic mice. Matrix Biol. 14, 181–190.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., Benfey, P.N., 2003. A gene expression map of the *Arabidopsis* root. Science 302, 1956–1960.

Bolz, J., Uziel, D., Muhlfriedel, S., Gullmar, A., Peuckert, C., Zarbalis, K., Wurst, W., Torii, M., Levitt, P., 2004. Multiple roles of ephrins during the formation of thalamocortical projections: maps and more. J. Neurobiol. 59, 82–94.

Carter, M.G., Hamatani, T., Sharov, A.A., Carmack, C.E., Qian, Y., Aiba, K., Ko, N.T., Dudekula, D.B., Brzoska, P.M., Hwang, S.S., Ko, M.S., 2003. In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. Genome Res. 13, 1011–1021.

Chua, S.S., Ma, Z., Ngan, E., Tsai, S.Y., 2004. Cdc25B as a steroid receptor coactivator. Vitam. Horm. 68, 231–256.

Churchill, G.A., 2002. Fundamentals of experimental design for cDNA microarrays. Nat. Genet. 32, 490–495.

Cochard, P., Paulin, D., 1984. Initial expression of neurofilaments and vimentin in the central and peripheral nervous system of the mouse embryo in vivo. J. Neurosci. 4, 2080–2094.

Dzierzak, E., Medvinsky, A., 1995. Mouse embryonic hematopoiesis. Trends Genet. 11, 359–366.

Echtay, K.S., Murphy, M.P., Smith, R.A., Talbot, D.A., Brand, M.D., 2002. Superoxide activates mitochondrial uncoupling protein 2 from the matrix side, studies using targeted antioxidants. J. Biol. Chem. 277, 47129–47135.

Edwards, M.A., Crandall, J.E., Wood, J.N., Tanaka, H., Yamamoto, M., 1989. Early axonal differentiation in mouse CNS delineated by an antibody recognizing extracted neurofilaments. Brain Res. Dev. Brain Res. 49, 185–204.

Furlong, E.E., Andersen, E.C., Null, B., White, K.P., Scott, M.P., 2001. Patterns of gene expression during *Drosophila* mesoderm development. Science 293, 1629–1633.

Grün, D., Wang, Y.-L., Langenberger, D., Gunsalus, K.C., Rajewsky, N., 2005. MicroRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. PLoS Comput. Biol. 1, 51–66.

Hamatani, T., Carter, M.G., Sharov, A.A., Ko, M.S., 2004. Dynamics of global gene expression changes during mouse preimplantation development. Dev. Cell 6, 117–131.

Hardman, M.J., Sisi, P., Banbury, D.N., Byrne, C., 1998. Patterned acquisition of skin barrier function during development. Development 125, 1541–1552.

Hengst, L., Reed, S.I., 1996. Translational control of p27Kip1 accumulation during the cell cycle. Science 271, 1861–1864.

Hill, A.A., Hunter, C.P., Tsung, B.T., Tucker-Kellogg, G., Brown, E.L., 2000. Genomic analysis of gene expression in *C. elegans*. Science 290, 809–812.

Huber, K., Bruhl, B., Guillemot, F., Olson, E.N., Ernsberger, U., Unsicker, K.,

2002. Development of chromaffin cells depends on MASH1 function. Development 129, 4729–4738.
Jacobsson, A., Stadler, U., Glotzer, M.A., Kozak, L.P., 1985. Mitochondrial uncoupling protein from mouse brown fat. Molecular cloning, genetic mapping, and mRNA expression. J. Biol. Chem. 260, 16250–16254.
Jaubert, J., Cheng, J., Segre, J.A., 2003. Ectopic expression of kruppel like factor 4 (Klf4) accelerates formation of the epidermal permeability barrier. Development 130, 2767–2777.
Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V., Kim, S.K., 2001. Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. 98, 218–223.
Johansson, K.A., Grapin-Botton, A., 2002. Development and diseases of the pancreas. Clin. Genet. 62, 14–23.
Julien, J.P., Meyer, D., Flavell, D., Hurst, J., Grosveld, F., 1986. Cloning and developmental expression of the murine neurofilament gene family. Brain Res. 387, 243–250.
Kageyama, R., Ishibashi, M., Takebayashi, K., Tomita, K., 1997. bHLH transcription factors and mammalian neuronal differentiation. Int. J. Biochem. Cell Biol. 29, 1389–1399.
Kaufman, M.H., Bard, J., 1999. The Anatomical Basis of Mouse Development. Academic Press, San Diego.
Kendziorski, C.M., Zhang, Y., Lan, H., Attie, A.D., 2003. The efficiency of pooling mRNA in microarray experiments. Biostatistics 4, 465–477.
Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N., Davidson, G.S., 2001. A gene expression map for *Caenorhabditis elegans*. Science 293, 2087–2092.
Ko, M.S., Kitchen, J.R., Wang, X., Threat, T.A., Hasegawa, A., Sun, T., Grahovac, M.J., Kargul, G.J., Lim, M.K., Cui, Y., Sano, Y., Tanaka, T., Liang, Y., Mason, S., Paonessa, P.D., Sauls, A.D., DePalma, G.E., Sharara, R., Rowe, L.B., Eppig, J., Morrell, C., Doi, H., 2000. Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development 127, 1737–1749.
Komatsu, N., Takata, M., Otsuki, N., Toyama, T., Ohka, R., Takehara, K., Saijoh, K., 2003. Expression and localization of tissue kallikrein mRNAs in human epidermis and appendages. J. Invest. Dermatol. 121, 542–549.
Koury, M.J., Sawyer, S.T., Brandt, S.J., 2002. New insights into erythropoiesis. Curr. Opin. Hematol. 9, 93–100.
Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., Rajewsky, N., 2005. Combinatorial microRNA target predictions. Nat. Genet. 37, 495–500.
Krishan, K., Morgan, M.J., Zhao, W., Dhoot, G.K., 2000. Slow troponin T mRNA in striated muscles is expressed in both cell type and developmental stage specific manner. J. Muscle Res. Cell Motil. 21, 527–536.
Ladomery, M., Dellaire, G., 2002. Multifunctional zinc finger proteins in development and disease. Ann. Hum. Genet. 66, 331–342.
Leung, C.L., Green, K.J., Liem, R.K., 2002. Plakins: a family of versatile cytolinker proteins. Trends Cell Biol. 12, 37–45.
Lu, M.F., Cheng, H.T., Kern, M.J., Potter, S.S., Tran, B., Diekwisch, T.G., Martin, J.F., 1999a. prx-1 functions cooperatively with another paired-related homeobox gene, prx-2, to maintain cell fates within the craniofacial mesenchyme. Development 126, 495–504.
Lu, M.F., Cheng, H.T., Lacy, A.R., Kern, M.J., Argao, E.A., Potter, S.S., Olson, E.N., Martin, J.F., 1999b. Paired-related homeobox genes cooperate in handplate and hindlimb zeugopod morphogenesis. Dev. Biol. 205, 145–157.
Mendelson, C.R., 2000. Role of transcription factors in fetal lung development and surfactant protein gene expression. Annu. Rev. Physiol. 62, 875–915.
Miki, R., Kadota, K., Bono, H., Mizuno, Y., Tomaru, Y., Carninci, P., Itoh, M., Shibata, K., Kawai, J., Konno, H., Watanabe, S., Sato, K., Tokusumi, Y., Kikuchi, N., Ishii, Y., Hamaguchi, Y., Nishizuka, I., Goto, H., Nitanda, H., Satomi, S., Yoshiki, A., Kusakabe, M., DeRisi, J.L., Eisen, M.B., Iyer, V.R., Brown, P.O., Muramatsu, M., Shimada, H., Okazaki, Y., Hayashizaki, Y., 2001. Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. Proc. Natl. Acad. Sci. U. S. A. 98, 2199–2204.
Moreno, N., Bachy, I., Retaux, S., Gonzalez, A., 2004. LIM-homeodomain genes as developmental and adult genetic markers of *Xenopus* forebrain functional subdivisions. J. Comp. Neurol. 472, 52–72.
Nedergaard, J., Golozoubova, V., Matthias, A., Asadi, A., Jacobsson, A., Cannon, B., 2001. UCP1: the only protein able to mediate adaptive non-shivering thermogenesis and metabolic inefficiency. Biochim. Biophys. Acta 1504, 82–106.
Ngan, E.S., Hashimoto, Y., Ma, Z.Q., Tsai, M.J., Tsai, S.Y., 2003. Over-expression of Cdc25B, an androgen receptor coactivator, in prostate cancer. Oncogene 22, 734–739.
Ogata, A., Shimizu, T., Abe, R., Shimizu, H., Sakai, M., 2004. Expression of c-maf and mafB genes in the skin during rat embryonic development. Acta Histochem. 106, 65–67.
Ohishi, K., Varnum-Finney, B., Bernstein, I.D., 2002. Delta-1 enhances marrow and thymus repopulating ability of human CD34(+)CD38(−) cord blood cells. J. Clin. Invest. 110, 1165–1174.
Okamoto, K., Li, H., Jensen, M.R., Zhang, T., Taya, Y., Thorgeirsson, S.S., Prives, C., 2002. Cyclin G recruits PP2A to dephosphorylate Mdm2. Mol. Cell 9, 761–771.
Paululat, A., Breuer, S., Renkawitz-Pohl, R., 1999. Determination and development of the larval muscle pattern in *Drosophila melanogaster*. Cell Tissue Res. 296, 151–160.
Peng, X., Wood, C.L., Blalock, E.M., Chen, K.C., Landfield, P.W., Stromberg, A.J., 2003. Statistical implications of pooling RNA samples for microarray experiments. BMC Bioinformatics 4, 26.
Poulsen, K., Jakobsen, B.K., Mikkelsen, B.M., Harmark, K., Nielsen, J.T., Hjorth, J.P., 1986. Coordination of murine parotid secretory protein and salivary amylase expression. EMBO J. 5, 1891–1896.
Rendl, M., Ban, J., Mrass, P., Mayer, C., Lengauer, B., Eckhart, L., Declerq, W., Tschachler, E., 2002. Caspase-14 expression by epidermal keratinocytes is regulated by retinoids in a differentiation-associated manner. J. Invest. Dermatol. 119, 1150–1155.
Ricquier, D., Bouillaud, F., 2000. The uncoupling protein homologues: UCP1, UCP2, UCP3, StUCP and AtUCP. Biochem. J. 345 (Pt 2), 161–179.
Robinson, M.L., Overbeek, P.A., 1996. Differential expression of alpha A- and alpha B-crystallin during murine ocular development. Invest. Ophthalmol. Visual Sci. 37, 2276–2284.
Russell, L., Forsdyke, D.R., 1991. A human putative lymphocyte G0/G1 switch gene containing a CpG-rich island encodes a small basic protein with the potential to be phosphorylated. DNA Cell Biol. 10, 581–591.
Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J., 2003. TM4: a free, open-source system for microarray data management and analysis. BioTechniques 34, 374–378.
Segre, J., 2003. Complex redundancy to build a simple epidermal permeability barrier. Curr. Opin. Cell Biol. 15, 776–782.
Shawlot, W., Behringer, R.R., 1995. Requirement for Lim1 in head-organizer function. Nature 374, 425–430.
Sherlock, G., 2000. Analysis of large-scale gene expression data. Curr. Opin. Immunol. 12, 201–205.
Shu, W., Jiang, Y.Q., Lu, M.M., Morrisey, E.E., 2002. Wnt7b regulates mesenchymal proliferation and vascular development in the lung. Development 129, 4831–4842.
Smith, T.H., Kachinsky, A.M., Miller, J.B., 1994. Somite subdomains, muscle cell origins, and the four muscle regulatory factor proteins. J. Cell Biol. 127, 95–105.
Stanley, W.C., Chandler, M.P., 2002. Energy metabolism in the normal and failing heart: potential for therapeutic interventions. Heart Fail. Rev. 7, 115–130.
Sundaresan, V., Mambetisaeva, E., Andrews, W., Annan, A., Knoll, B., Tear, G., Bannister, L., 2004. Dynamic expression patterns of Robo (Robo1 and Robo2) in the developing murine central nervous system. J. Comp. Neurol. 468, 467–481.
Tabibiazar, R., Wagner, R.A., Liao, A., Quertermous, T., 2003. Transcriptional profiling of the heart reveals chamber-specific gene expression patterns. Circ. Res. 93, 1193–1201.
Tabibiazar, R., Wagner, R.A., Ashley, E.A., King, J.Y., Ferrara, R., Spin, J.M.,

Janssen, J.J., Narasimhan, B., Tibshirani, R., Tsao, P.S., Efron, B., Quertermous, T., 2005a. Signature patterns of gene expression in mouse atherosclerosis and their correlation to human coronary disease. Physiol. Genomics 22 (2), 213–226.

Tabibiazar, R., Wagner, R.A., Spin, J.M., Ashley, E.A., Narasimhan, B., Rubin, E.M., Efron, B., Tsao, P.S., Tibshirani, R., Quertermous, T., 2005b. Mouse strain-specific differences in vascular wall gene expression and their relationship to vascular disease. Arterioscler. Thromb. Vasc. Biol. 25, 302–308.

Tesfaigzi, J., Carlson, D.M., 1999. Expression, regulation, and function of the SPR family of proteins, a review. Cell Biochem. Biophys. 30, 243–265.

Tohda, S., Murata-Ohsawa, M., Sakano, S., Nara, N., 2003. Notch ligands, Delta-1 and Delta-4 suppress the self-renewal capacity and long-term growth of two myeloblastic leukemia cell lines. Int. J. Oncol. 22, 1073–1079.

Tusher, V.G., Tibshirani, R., Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U. S. A. 98, 5116–5121.

Van Limpt, V.A., Chan, A.J., Van Sluis, P.G., Caron, H.N., Van Noesel, C.J., Versteeg, R., 2003. High delta-like 1 expression in a subset of neuroblastoma cell lines corresponds to a differentiated chromaffin cell type. Int. J. Cancer 105, 61–69.

Villarroya, F., Brun, S., Giralt, M., Camara, Y., Solanes, G., Iglesias, R., 2001. Gene expression of leptin and uncoupling proteins: molecular end-points of fetal development. Biochem. Soc. Trans. 29, 76–80.

Wagner, R.A., Tabibiazar, R., Powers, J., Bernstein, D., Quertermous, T., 2004. Genome-wide expression profiling of a cardiac pressure overload model identifies major metabolic and signaling pathway responses. J. Mol. Cell Cardiol. 37, 1159–1170.

Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., Zernicka-Goetz, M., 2004. A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Dev. Cell 6, 133–144.

Whitsett, J.A., Weaver, T.E., 2002. Hydrophobic surfactant proteins in lung function and disease. N. Engl. J. Med. 347, 2141–2148.

Yang, H., Lu, M.M., Zhang, L., Whitsett, J.A., Morrisey, E.E., 2002. GATA6 regulates differentiation of distal lung epithelium. Development 129, 2233–2246.

Yoon, J.K., Olson, E.N., Arnold, H.H., Wold, B.J., 1997. Different MRF4 knockout alleles differentially disrupt Myf-5 expression: *cis*-regulatory interactions at the MRF4/Myf-5 locus. Dev. Biol. 188, 349–362.

Zeeberg, B.R., Feng, W., Wang, G., Wang, M.D., Fojo, A.T., Sunshine, M., Narasimhan, S., Kane, D.W., Reinhold, W.C., Lababidi, S., Bussey, K.J., Riss, J., Barrett, J.C., Weinstein, J.N., 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol. 4, R28.

Zeeuwen, P.L., Van Vlijmen-Willems, I.M., Jansen, B.J., Sotiropoulou, G., Curfs, J.H., Meis, J.F., Janssen, J.J., Van Ruissen, F., Schalkwijk, J., 2001. Cystatin M/E expression is restricted to differentiated epidermal keratinocytes and sweat glands: a new skin-specific proteinase inhibitor that is a target for cross-linking by transglutaminase. J. Invest. Dermatol. 116, 693–701.

Zeng, F., Baldwin, D.A., Schultz, R.M., 2004. Transcript profiling during preimplantation mouse development. Dev. Biol. 272, 483–496.

Zhang, J., 2002. Myocardial energetics in cardiac hypertrophy. Clin. Exp. Pharmacol. Physiol. 29, 351–359.

Zhou, L., Lim, L., Costa, R.H., Whitsett, J.A., 1996. Thyroid transcription factor-1, hepatocyte nuclear factor-3beta, surfactant protein B, C, and Clara cell secretory protein in developing mouse lung. J. Histochem. Cytochem. 44, 1183–1193.